GENETIC ANALYSIS OF SOUTH AMERICAN EASTERN EQUINE ENCEPHALOMYELITIS VIRUSES ISOLATED FROM MOSQUITOES COLLECTED IN THE AMAZON BASIN REGION OF PERU

JOHN P. KONDIG, MICHAEL J. TURELL,* JOHN S. LEE, MONICA L. O'GUINN, AND LEONARD P. WASIELOSKI JR

Diagnostics Systems Division and Virology Division, U.S. Army Medical Research Institute of Infectious Diseases, Fort Detrick, Maryland

Abstract. Identifying viral isolates from field-collected mosquitoes can be difficult and time-consuming, particularly in regions of the world where numerous closely related viruses are co-circulating (e.g., the Amazon Basin region of Peru). The use of molecular techniques may provide rapid and efficient methods for identifying these viruses in the laboratory. Therefore, we determined the complete nucleotide sequence of two South American eastern equine encephalomyelitis viruses (EEEVs): one member from the Peru-Brazil (Lineage II) clade and one member from the Argentina-Panama (Lineage III) clade. In addition, we determined the nucleotide sequence for the nonstructural P3 protein (nsP3) and envelope 2 (E2) protein genes of 36 additional isolates of EEEV from mosquitoes captured in Peru between 1996 and 2001. The 38 isolates were evenly distributed between lineages II and III virus groupings. However, analysis of the nsP3 gene for lineage III strongly suggested that the 19 isolates from this lineage could be divided into two sub-clades, designated as lineages III and IIIA. Compared with North American EEEV (lineage I, GA97 strain), we found that the length of the nsP3 gene was shorter in the strains isolated from South America. A total of 60 nucleotides was deleted in lineage II, 69 in lineage III, and 72 in lineage IIIA. On the basis of the sequences we determined for South American EEEVs and those for other viruses detected in the same area, we developed a series of primers for characterizing these viruses.

INTRODUCTION

Identifying arthropod-borne viruses present in field-collected specimens can be problematic and may require extensive amounts of time, particularly when the samples are from regions of the world where diverse assortments of arthropods and viruses co-circulate (e.g., the Amazon Basin region of Peru in South America). For traditional virus identification, antibody must be prepared against the new virus isolate and then that virus and antibody combination must be tested against other known, closely related viruses and antibody preparations to determine the relative relationship of the new virus isolate with the known viruses and antibodies used in the detection-characterization assay. Not only is this time-consuming and expensive, but if the actual virus is not included in the characterization assay, then the virus isolate might mistakenly be declared a new virus (e.g., Zinga is really Rift Valley fever virus).[1] The advent of molecular diagnostic tools has allowed the development of rapid and specific assays for numerous viruses across many virus families. Once a virus has been isolated and generically characterized by using a broadly cross-reactive test (e.g., an immunofluorescence assay [IFA] using polyclonal sera), a series of polymerase chain reaction (PCR) primers can be designed and used to confirm the identity of the virus. In many cases, the sequence determined from the PCR amplicon can be used for determining the genetic relationship of the new virus isolate with other viruses isolated from the same region or from different regions of the world. This in turn may provide clues as to virus maintenance mechanisms and the sources of virus infections in the local population.

During our examination of more than 160 virus isolates from the Amazon Basin region of Peru,[2] we developed an approach that allowed us to rapidly identify members of the *Alphavirus* and *Flavivirus* genera and to obtain sequence information that allowed us to compare directly the relationship between these viruses and to other viruses isolated in the same region. Over the five years of the study, eastern equine encephalomyelitis virus (EEEV) was the most commonly isolated virus. Infection with EEEV in either humans or equines can result in a serious, often fatal disease[3] and is an important public health concern in North and South America. Although EEEV is transmitted by mosquitoes in North America (primarily east of the Mississippi River) and in South America, the strains of virus circulating in these two regions differ significantly.[4] Eastern equine encephalomyelitis virus can be separated into four subtypes, based on genetic information.[4,5] These include one subtype found in North America, Lineage I (isolates from North America), and three subtypes found in South America, Lineage II (Brazil-Peru) (isolates found in Brazil, Guatemala, and Peru), Lineage III (Argentina-Panama) (isolates found in Argentina, Brazil, Colombia, Ecuador, Guiana, Panama, Peru, Trinidad, and Venezuela), and Lineage IV (a single isolate from Brazil).[5] This report will examine the genetic relationship of 38 isolates of EEEV made from mosquitoes captured in the Amazon Basin region of Peru from 1996 until 2001 and will compare these results with published data for other South American isolates of EEEV.

METHODS

Virus isolation and identification. Table 1 lists the South America EEEV isolates that were identified and for which the sequences of the nonstructural protein 3 (nsP3) and envelope 2 (E2) genes were determined. Viruses were isolated from mosquitoes collected at several sites near Iquitos, Loreto Department, in the Amazon Basin in northeastern Peru. Iquitos (population approximately 300,000) is approximately 125 meters above sea level and is bordered by the Amazon, Itaya, and Nanay Rivers (3°51′S, 73°13′W). Meth-

* Address correspondence to Michael J. Turell, Department of Cell Biology, Virology Division, U.S. Army Medical Research Institute of Infectious Diseases, 1425 Porter Street, Fort Detrick, MD 21702-5011. E-mail: michael.turell@det.amedd.army.mil

Report Documentation Page

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE **1 MAR 2007**	2. REPORT TYPE **N/A**	3. DATES COVERED **-**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Genetic analysis of South American eastern equine encephalitis viruses isolated from mosquitoes collected in the Amazon Basin region of Peru. American Journal of Tropical Medicine and Hygiene 76:408-416	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)	5d. PROJECT NUMBER
Kondig, JP Turell, MJ Lee, JS O'Guinn, ML Wasieloski, LP Jr	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **United States Army Medical Research Institute of Infectious Diseases, Fort Detrick, MD**	8. PERFORMING ORGANIZATION REPORT NUMBER **RPP-06-044**

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Identifying viral isolates from field-collected mosquitoes can be difficult and time-consuming, particularly in regions of the world where numerous closely related viruses are co-circulating (e.g., the Amazon Basin region of Peru). The use of molecular techniques may provide rapid and efficient methods for identifying these viruses in the laboratory. Therefore, we determined the complete nucleotide sequence of two South American eastern equine encephalomyelitis viruses (EEEVs): one member from the Peru-Brazil (Lineage II) clade and one member from the Argentina-Panama (Lineage III) clade. In addition, we determined the nucleotide sequence for the nonstructural P3 protein (nsP3) and envelope 2 (E2) protein genes of 36 additional isolates of EEEV from mosquitoes captured in Peru between 1996 and 2001. The 38 isolates were evenly distributed between lineages II and III virus groupings. However, analysis of the nsP3 gene for lineage III strongly suggested that the 19 isolates from this lineage could be divided into two sub-clades, designated as lineages III and IIIA. Compared with North American EEEV (lineage I, GA97 strain), we found that the length of the nsP3 gene was shorter in the strains isolated from South America. A total of 60 nucleotides was deleted in lineage II, 69 in lineage III, and 72 in lineage IIIA. On the basis of the sequences we determined for South American EEEVs and those for other viruses detected in the same area, we developed a series of primers for characterizing these viruses.

15. SUBJECT TERMS
eastern equine encephalitis virus, EEE, genetic analysis, mosquito vector, Peru, isolation

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT **SAR**	18. NUMBER OF PAGES **9**	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**			

TABLE 1
Strains of eastern equine encephalitis virus examined in the study

Isolate no.	Mosquito species*	Location collected†	Date collected	Lineage	Accession no.‡	
					E2	nsP3
PE-0.0155	*Cx. pedroi*	PA Forest	8/2/96	III	DQ241304‡	
PE-1.0643	*Cx. pedroi*	PA Forest	9/20/96	III	DQ280382	DQ307803
PE-2.0010	*Cx. pedroi*	PA Forest	10/23/95	II	DQ280417	DQ307838
PE-3.0041	*Cx. pedroi*	PA Forest	12/3/96	III	DQ280384	DQ307805
PE-3.0391	*Cx. pedroi*	PA Forest	12/5/96	III	DQ280385	DQ307806
PE-3.0803	*Cx. pedroi*	PA Forest	12/10/96	III	DQ280386	DQ307807
PE-3.0815§	*Cx. pedroi*	PA Forest	12/10/96	II	DQ241303§	
PE-3.0869	*Cx. pedroi*	PA Forest	12/10/96	III	DQ280387	DQ307808
PE-4.0661	*Cx. pedroi*	Casa Juan	11/22/97	II	DQ280416	DQ307837
PE-4.0775	*Cx. pedroi*	PA Forest	1/26/97	III	DQ280388	DQ307809
PE-4.0807	*Cx. pedroi*	PA Forest	1/26/97	II	DQ280415	DQ307836
PE-4.0808	*Cx. pedroi*	PA Forest	1/26/97	II	DQ280414	DQ307835
PE-5.0151	*Cx. (Mel.)* spp.	PA Forest	2/26/97	III	DQ280389	DQ307810
PE-5.0183	*Ps. albigenu*	PA Forest	2/26/97	III	DQ280390	DQ307811
PE-5.0519	*Oc. fulvus*	PA Forest	2/26/97	III	DQ280391	DQ307812
PE-10.0046	*Cx. pedroi*	PA Forest	8/13/97	II	DQ280413	DQ307834
PE-10.0170	*Cx. pedroi*	PA Forest	8/25/97	II	DQ280412	DQ307833
PE-11.0042	*Cx. pedroi*	PA Forest	9/18/97	III	DQ280392	DQ307813
PE-11.0207	*Cx. pedroi*	Mixed	9/22/97	II	DQ280411	DQ307832
PE-11.0331	*Cx. pedroi*	PA Forest	9/29/97	III	DQ280393	DQ307814
PE-11.0352	*Cx. pedroi*	PA Forest	9/29/97	III	DQ280394	DQ307815
PE-15.0058	*Cs. pedroi*	PA Forest	8/98	II	DQ280410	DQ307831
PE-16.0050	*Cx. pedroi*	PA Forest	9/23/98	III	DQ280395	DQ307816
PE-16.0140	*Cx. pedroi*	Otorongo	10/4/98	III	DQ280396	DQ307817
PE-1.0999	Hamster	PA area	2/28/98	III	DQ280383	DQ307804
PE-17.0547	*Cx. pedroi*	Otorongo	12/8/98	III	DQ280397	DQ307818
PE-18.0140	*Cx. pedroi*	PA Forest	2/17/99	II	DQ280409	DQ307830
PE-18.0169	*Cx. pedroi*	PA Forest	2/17/99	II	DQ280408	DQ307829
PE-18.1150	*Cx. pedroi*	Eng	2/25/99	II	DQ280407	DQ307828
PE-22.0110	*Cx. pedroi*	PA Forest	2/9/00	III	DQ280398	DQ307819
PE-22.0263	*Cx. pedroi*	PA Forest	10/2/00	II	DQ280406	DQ307827
PE-22.0285	*Cx. gnomatos*	PA Forest	10/2/00	II	DQ280405	DQ307826
PE-22.0526	*Cx. pedroi*	ACEER	2/00	II	DQ280404	DQ307825
PE-22.0534	*Cx. pedroi*	ACEER	2/15/00	II	DQ280403	DQ307824
PE-22.0552	*Cx. pedroi*	ACEER	2/15/00	III	DQ280399	DQ307820
PE-22.0678	*Cx. pedroi*	Otorongo	2/20/00	II	DQ280402	DQ307823
PE-24.0111	*Cx. pedroi*	ACEER	9/6/00	II	DQ280401	DQ307822
PE-24.0132	*Cx. pedroi*	ACEER	9/6/00	II	DQ280400	DQ307821

* *Cx.* = *Culex; Mel.* = *Melanoconian; Ps.* = *Psorophora; Oc.* = *Ochlerotatus.*
† Locations: PA Forest = Puerto Almendras Forest (approximately 20 km west-southwest of Iquitos); Casa Juan = private home about 4 km northeast of PA Forest; Otorongo - Peruvian military training base 15 km southwest of PA Forest, Eng = Peruvian army engineering base located about 20 km southwest of PA Forest; ACEER = Amazon Center for Environmental Education and Research located on the Napo River about 180 km north of PA Forest.
‡ Accession numbers for GenBank for the sequence of the envelope 2 (E2) and nonstructural protein 3 (nsP3) regions.
§ Accession numbers for GenBank for the complete genome sequence.

ods for mosquito collection, virus isolation, and initial identification were as previously described.[2] Briefly, mosquitoes were captured in dry ice-baited miniature light traps (John W. Hock Co., Gainesville, FL), sorted to species, placed in pools, and frozen at −70°C until tested for infectious virus by plaque assay on Vero (African green monkey kidney) cells. Viruses that grew rapidly and produced plaques by day 2 were broadly categorized as alphaviruses (however, this group also included some of the more rapidly growing bunyaviruses), and viruses that grew more slowly and produced plaques by day 5 were broadly categorized as flaviviruses (however, this group also included some of the more slowly growing bunyaviruses). Viruses were amplified in Vero cell cultures and were initially screened by IFA for reactivity against *Alphavirus* and *Flavivirus* genus-specific monoclonal antibodies (U.S. Army Medical Research Institute of Infectious Diseases, Fort Detrick, MD). Follow-up IFA tests were performed using available antisera to complex or virus-specific members according to standard procedures.[6] Based on these results, the identification of viral isolates as members of the genera *Fla-*

vivirus and *Alphavirus* was confirmed by reverse transcription–PCR (RT-PCR) and sequencing of the PCR amplicons.

RNA extraction and PCR amplification. Viral RNA was isolated from virus-infected cell culture supernatant using TRIzol-LS Reagent (Invitrogen, Carlsbad, CA) according to the manufacturer's protocol. The viral RNA was converted into cDNA using SuperScript™ II and primed with either oligo-dT or with random hexamers according to the manufacturer's instructions (Invitrogen). These cDNAs served as templates in subsequent PCRs containing virus-specific oligonucleotide primers (Table 2). The PCR amplifications were typically conducted in a PerkinElmer 2400 thermocycler (PerkinElmer Life and Analytical Sciences, Inc. Boston, MA,) in a total volume of 36 μL that contained 30 μL of high-fidelity PCR supermix (Invitrogen), 100 ng of cDNA template, and 50 pmol of each primer. The PCR amplicons were purified by using a QIAquick PCR purification kit (Qiagen, Valencia, CA) and 8 μL of the purified amplicon was added directly to the sequencing reaction.

To confirm the preliminary IFA identifications, cDNAs

TABLE 2
Primers used in the studies

Name	Use	Nucleotide sequence (5′ → 3′)	Reference
		Genus-specific primers	
0092-F	Alphavirus	GATGAAATCNGGVATGTT	12
0091-R	Alphavirus	ATTCAGGTTAGCCGTAGA	12
MA-F	Flavivirus	CATGATGGGRAARAGRGARRAG	13
cFD2-R	Flavivirus	GTGTCCCAGCCGGCGGTGTCATCAGC	13

Primers used to amplify the EEEV nsP3 and E2 genes*

| Primer | Lineage† | | Nucleotide sequence (5′ → 3′) | Gene |
	II	III		
nsP2A-F	3,554	NA	TAGGAACCCCAATTTCCG	nsP3
nsP2B-F	3,725	NA	AAGACCATGCCATTCACCACA	nsP3
nsP2F	3,931	3,931	GGCAAGGAYAATGGKAAC	nsP3
nsP4R	5,765	5,765	TCGAGGCGCGGGGCGTAG	nsP3
nsP4A-R	NA	5,798	CATAGTTGTAGCTTTCTCTGCAG	nsP3
nsP4B-R	6,244	NA	CCGCTAAAACGTTCTGAA	nsP3
CAP-AF	8,036	8,030	GATGTTCCACAATGTATG	E2
CAP-BF	8,274	8,268	GGAACCAGAAAGGRGTTA	E2
E1A-R	9,986	9,980	GGATCCCCACCTTGTTCGGCA	E2
E1B-R	10,275	10,269	CGTTCAATGTACRCTTCG	E2

* EEEV = eastern equine encephalomyelitis virus; nsP3 = nonstructural protein 3; E2 = envelope 2; NA = not applicable.
† Numbers correspond to complete genome sequence of either Lineage II or III.

were tested by PCR using *Alphavirus or Flavivirus* genus-specific primers (Table 2) and then tested in a separate PCR assay using South American EEEV-specific primers (Table 2). To further characterize the EEEV isolates, we PCR-amplified the entire nsP3 and E2 genes, using gene-specific primers (Table 2) that were based on the complete genome sequence of individual Brazil-Peru (Lineage II) and Argentina-Panama (Lineage III) South American EEEV published in GenBank. Most of the nsP3 amplicons were obtained using the primers nsP2F and nsP4R. The nsP3 amplicons for the remaining isolates were obtained using either primers nsP2A-F or nsP2B-F paired with primer nsP4B-R, or primers nsP2B-F or nsP2F paired with primer nsP4A-R. Primers CAP-AF and E1A-R or CAP-BF and E1B-R were used to amplify the E2 gene.

Sequence analysis. Sequencing was performed with Big Dye™ (Applied Biosystem, Inc., Foster City, CA) reagents according to the manufacturer's instructions on an Applied Biosystems 3100 ABI PRISM automated DNA sequencer (Applied Biosystems) and the sequence data were analyzed by the programs contained in the Lasergene suite of programs (Lasergene analysis software; DNASTAR, Inc., Madison, WI). Consensus sequences were determined by the SeqMan program (DNASTAR, Inc.). Alignment analysis of the consensus sequences were performed with the MegaAlign program (DNASTAR) using the default settings (gap penalty = 15, gap length penalty = 6.66, delay divergent sequences = % 30, DNA transition weight = 0.50, and DNA matrix ClustalW) of the Clustal W method.[7] The current version of Clustal W both aligns the sequences and also produces phylogenetic trees by the neighbor-joining method with the Kimura two-parameter distance formula as the default setting.[8,9] In addition, we used the Clustal W program to calculate bootstrap values with a default setting of 1,000 trials (iterations) and a seed value of 111. To reduce clutter in the figures, we only present bootstrap values ≥ 75 at the respective nodes. If bootstrap values were < 75, no value was indicated.

RESULTS

Using a two-step procedure, IFA and/or plaquing behavior to make a preliminary group identification, followed by RT-PCR with specific primers, we were able to identify rapidly > 75 of the virus isolations made from mosquitoes captured in the Amazon Basin region of Peru. These included members of the genus *Alphavirus* (EEE, Una, Venezuelan equine encephalomyelitis, and western equine encephalomyelitis [WEE] viruses) and the genus *Flavivirus* (Ilheus and St. Louis encephalitis viruses).[2] The genus- and virus-specific primers used are listed in Table 2. The genus-specific primers detected various members of that genus, but not members of other genera, and the EEEV primers detected South American EEEV.

The isolation of EEEV from 37 pools of mosquitoes and one hamster collected during this five-year study allowed us to examine the diversity of EEEV strains circulating in this region. Interestingly, 33 of the 37 EEEV isolations from mosquitoes were made from *Culex* (*Melanoconion*) *pedroi*. In addition, this species is a competent laboratory vector of EEEV (Turell MJ, unpublished data), indicating that this species is the principal vector of EEEV in this region of Peru. Phylogenetic analysis of the 38 isolates showed that they equally represented lineage II and III viruses (Table 1). The even distribution of the two lineages remained when the EEEV strains were compared by season; seven isolations from each lineage were made between August and September and 12 isolations from each lineage between December and February. Because mosquitoes infected with either a lineage II or III virus were co-captured in a single light trap on January 26, 1996 and on December 10, 1996, viruses in both clades were co-circulating in nature (Table 2).

Phylogenetic relationships for both the nsP3 and E2 genes for lineage II and III are shown in Figures 1 and 2. Evaluation of nsP3 and E2 lineage II genes showed a distinct group composed of six isolates (PE-4.0661, PE-10.0170, PE-11.0207, PE-18.1150, PE-22.0263, and PE-22.0678) (Figures 3A and 4A).
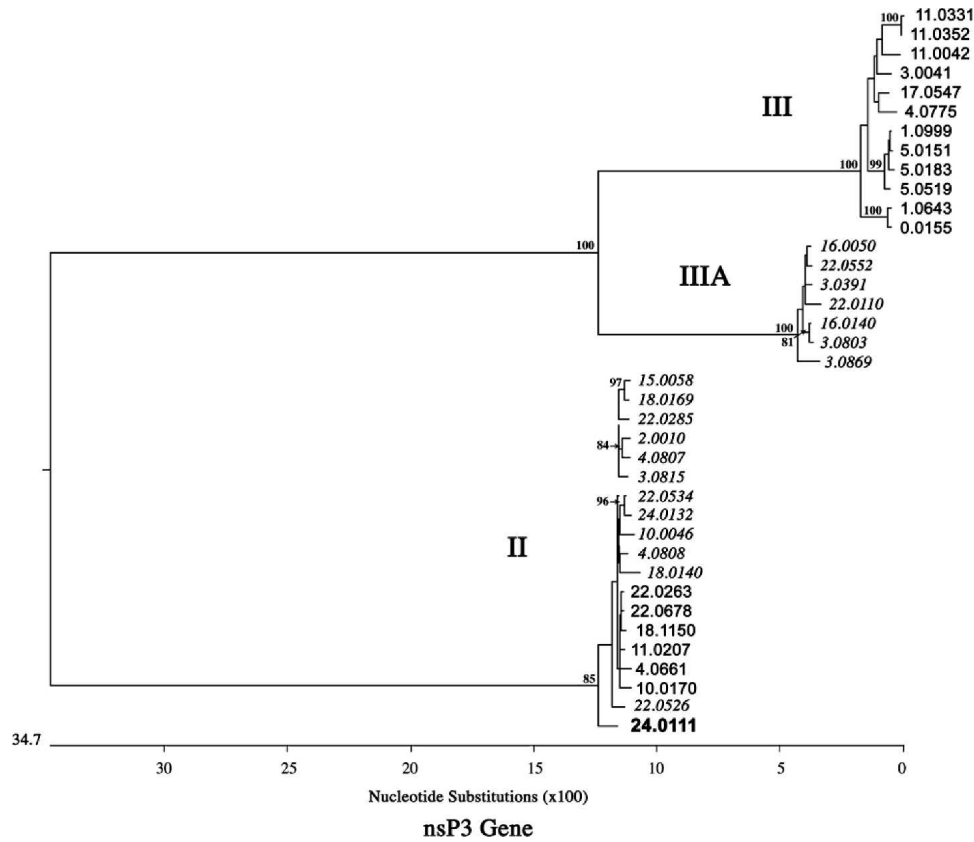
FIGURE 1. Phylogenetic tree of South American eastern equine encephalomyelitis virus isolates generated from a complete nonstructural protein 3 (nsP3) gene sequence for all viruses tested. The viruses in clade III and a subclade in lineage II are shown in regular type, and the remaining viruses in these two lineages are shown in italics. The single outlier in lineage II, PE-24.0111, is shown in **bold**. Bootstrap values ≥ 75 are shown at nodes.
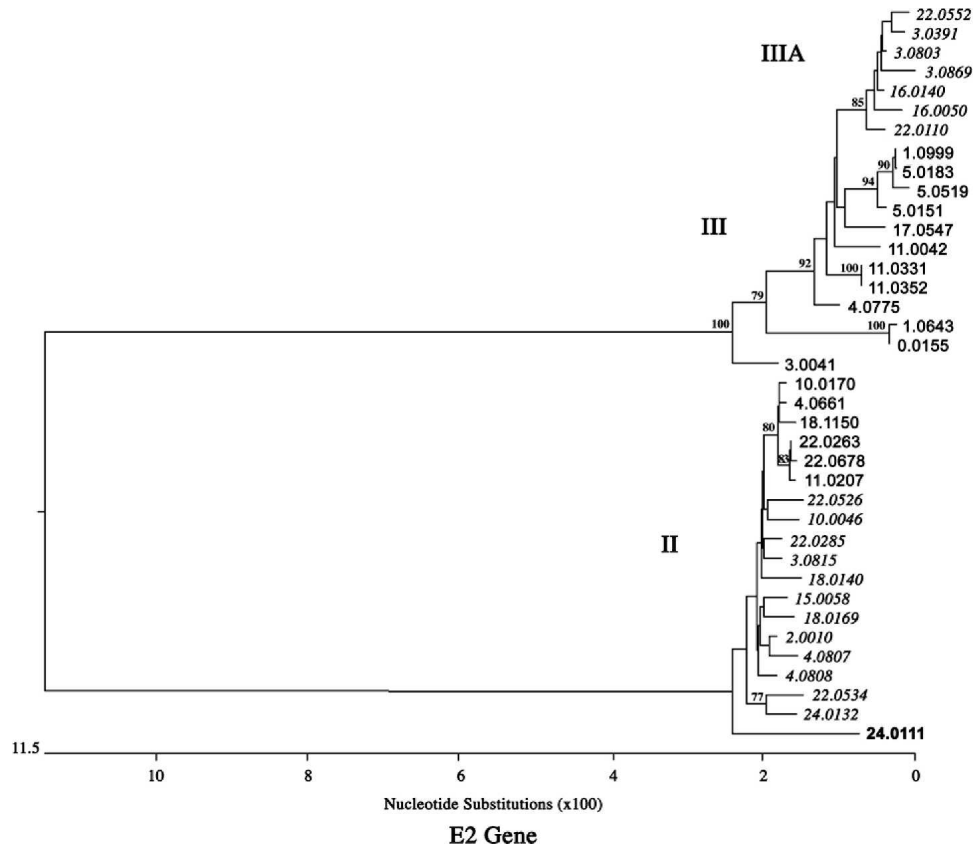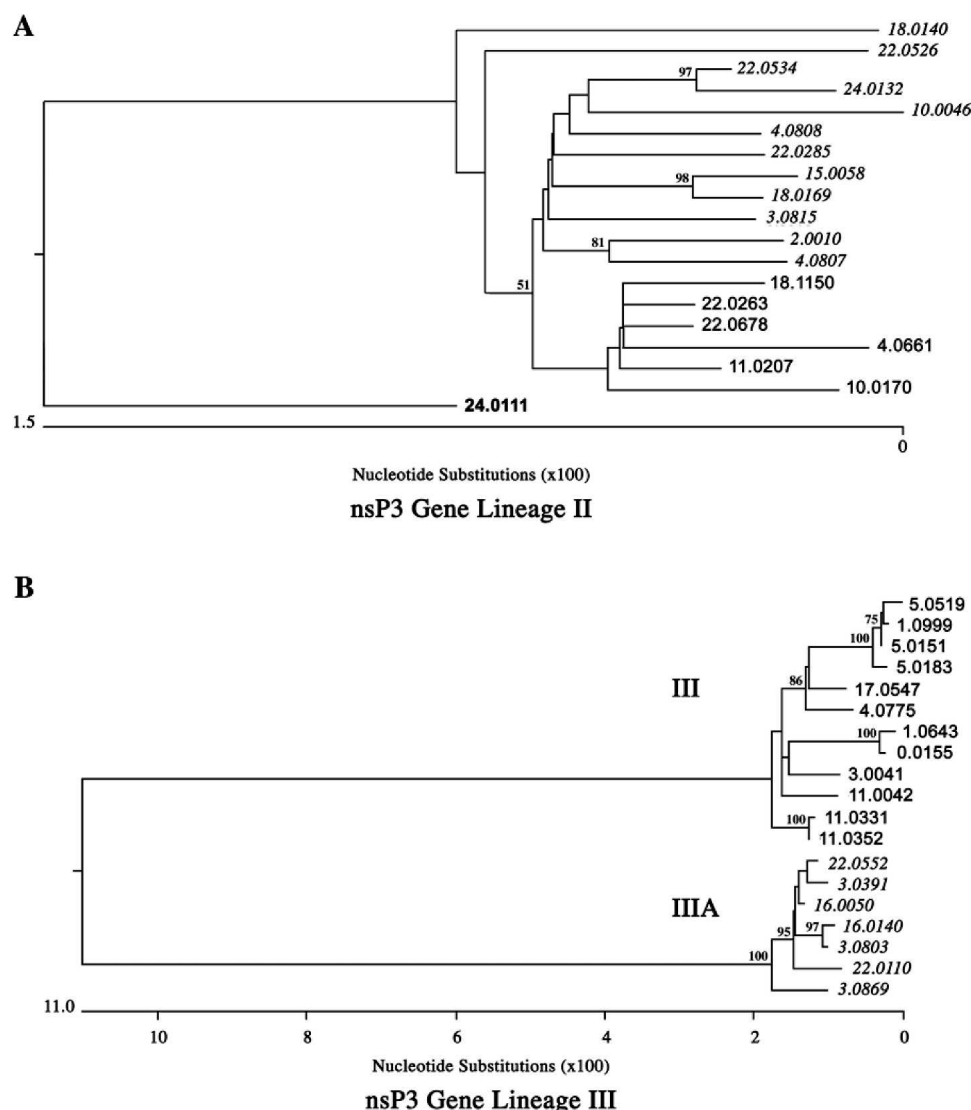

FIGURE 2. Phylogenetic tree of South American eastern equine encephalomyelitis virus isolates generated from complete envelope 2 (E2) protein gene sequences. The viruses in clade III and a subclade in lineage II are shown in regular type, and the remaining viruses in these two lineages are shown in italics. The single outlier in lineage II, PE-24.0111, is shown in **bold**. Bootstrap values ≥ 75 are shown at nodes.

FIGURE 3.   **A**, Expanded view of the nonstructural protein 3 (nsP3) genes for viruses in lineage II from Figure 1. Bootstrap values ≥ 75 are shown at nodes. **B**, Expanded view of the nsP3 genes for viruses in lineage III from Figure 1. Bootstrap values ≥ 75 are shown at nodes.

Although, the bootstrap values were only 51 (nsP3) and 24 (E2) for this group being distinct, the fact that these same six viruses formed a separate group based on both the nsP3 and E2 genes indicates that despite the low bootstrap values, the grouping may be real. The six isolates shared common nucleotide substitutions not found within the nsP3 and E2 genes of the other isolates. The nucleotide substitutions represented silent changes except at nucleotide position 9356 (genomic numbering) of isolate PE 3.0815, which resulted in a conservative change between valine and isoleucine. A single isolate, PE-24.0111, was distinct from all of the other isolates in both nsP3 and E2 analyses. Interestingly, PE-24.0111 and another clade II virus were isolated from mosquitoes caught in the same location on the same night (Table 2), indicating that distinct strains were co-circulating.

Evaluation of the nsP3 and E2 genes from lineage III showed two discrete subclades based on both the nsP3 and E2 genes (Figures 3B and 4B). The seven isolates (PE-3.0391, PE-3.0803, PE-3.0869, PE-16.0050, PE-16.0140, PE-22.0110, and PE-22.0552) formed a distinct subclade within the nsP3 gene alignment of lineage III with the Clustal W program (bootstrap value = 100). These seven isolates shared three unique characteristics in the nsP3 gene: 1) a deletion of three nucleotides at nucleotides 5299–5301 (PE-0.0155 genome numbering) that resulted in the deletion of a leucine residue; 2) a mutation at nucleotide 5296 resulting in either a proline or a serine; and 3) a mutation at nucleotide 5477 resulting in either a isoleucine or a threonine. All other shared nucleotide substitutions resulted in silent changes. Examination of the E2 gene from the seven nsP3-related isolates indicated that they also formed a separate subclade based on the E2 gene. However, this subclade was not as distinct (bootstrap = 92) because they contained shared changes at only two positions, nucleotide 8920 and nucleotide 9124 (PE-0.0155 genome numbering), but only at nucleotide 9124 was there an amino acid change, between a glutamine and a histidine (Figure 4B). There were three additional changes, but these were not all shared and unique to the subclade. An additional four isolates (PE-1.0999, PE-5.0151, PE-5.0183, and PE-5.0519) in lineage III appeared to form a separate subclade, but in all cases, the

A

Nucleotide Substitutions (x100)

E2 Gene Lineage II

B

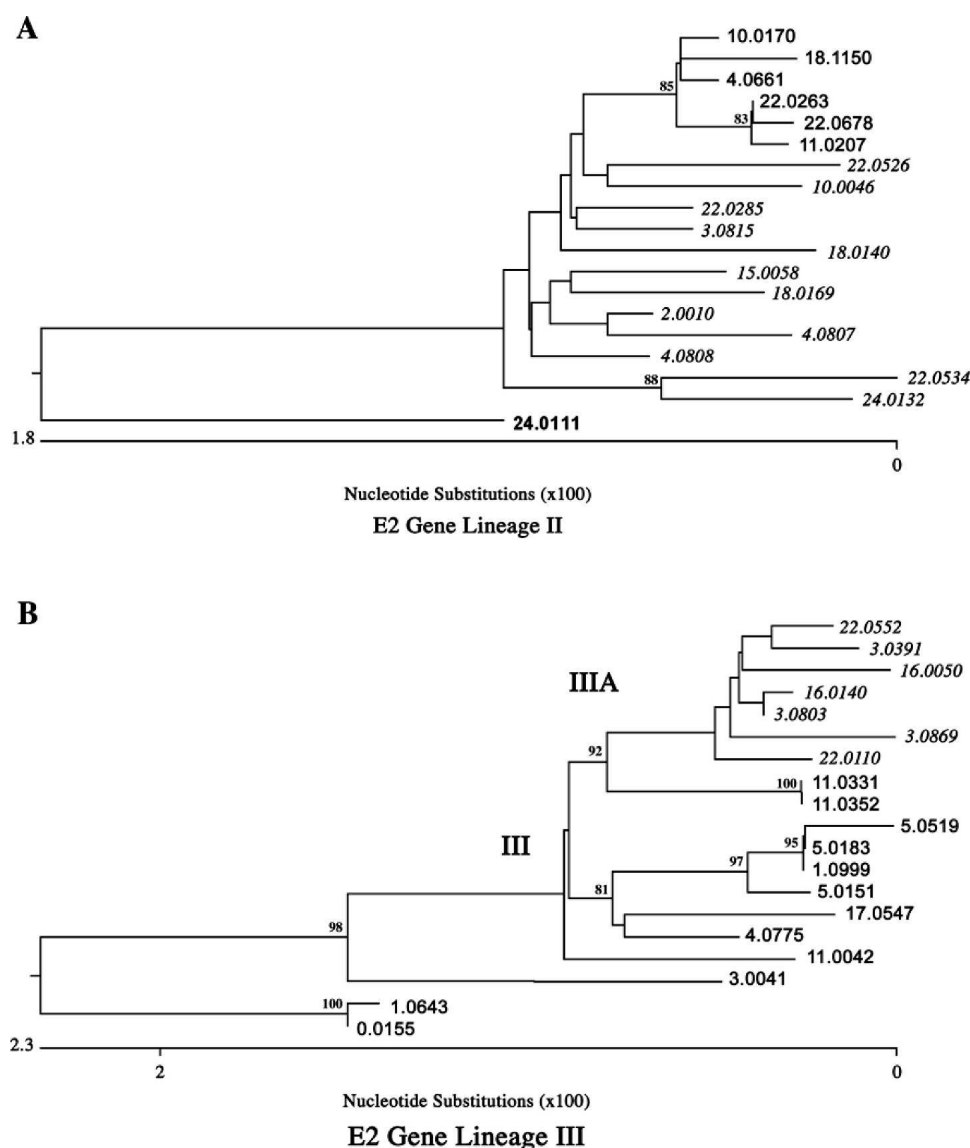Nucleotide Substitutions (x100)

E2 Gene Lineage III

FIGURE 4.   **A**, Expanded view of the E2 genes for viruses in lineage II from Figure 3. Bootstrap values ≥ 75 are shown at nodes. **B**, Expanded view of the E2 genes for viruses in lineage III from Figure 2. Bootstrap values ≥ 75 are shown at nodes.

shared nucleotide changes resulted in silent mutations in the nsP3 and E2 genes. Overall, 80–90% of the nucleotide changes resulted in silent substitutions, depending on the gene analyzed.

One characteristic shared by both lineage II and III viruses that differed from lineage I viruses (i.e., GA97) was the length of the nsP3 gene. A total of 60 nucleotides was deleted in lineage II, 69 nucleotides in lineage III, and 72 nucleotides in lineage IIIA (our designation) (Table 3). The distribution of deletions was consistent among all isolates of a specific lineage relative to lineage I. In all cases, the deletions/insertions-substitutions occurred near the 3′ end of the nsP3 gene, after nucleotide 4999 (GA97 genome numbering).

TABLE 3
Length of the nsP3 and E2 genes in selected clades of eastern equine encephalitis virus*

| Lineage | Clade† | Nucleotide length | |
		nsP3	E2
I	GA97	1,677	1,260
II	PB	1,617	1,260
III	PA	1,608	1,260
IIIA	PA	1,605	1,260

* nsP3 = nonstructural protein 3; E2 = envelope 2.
† GA97 = Georgia, North America; PB = Peru-Brazil; PA = Panama-Argentina.

TABLE 4
Relative identity of lineage I, II, and III viruses*

| Gene | Lineages compared† | | |
	I vs. II	II vs. III	I vs. III
Entire genome	66.4	81.8	71.6
E2 gene	67.1	82.1	66.1
nsP3 gene	61.9	67.5	62.8

* Genetic sequences were aligned and compared by the Martinez-NW Method. It uses two alignment methods in succession. Regions of perfect match are identified as described by Martinez[14] and the Needleman-Wunsch method[15] was then used to optimize the fit in between perfect matches. E2 = envelope 2; nsP3 = nonstructural protein 3.
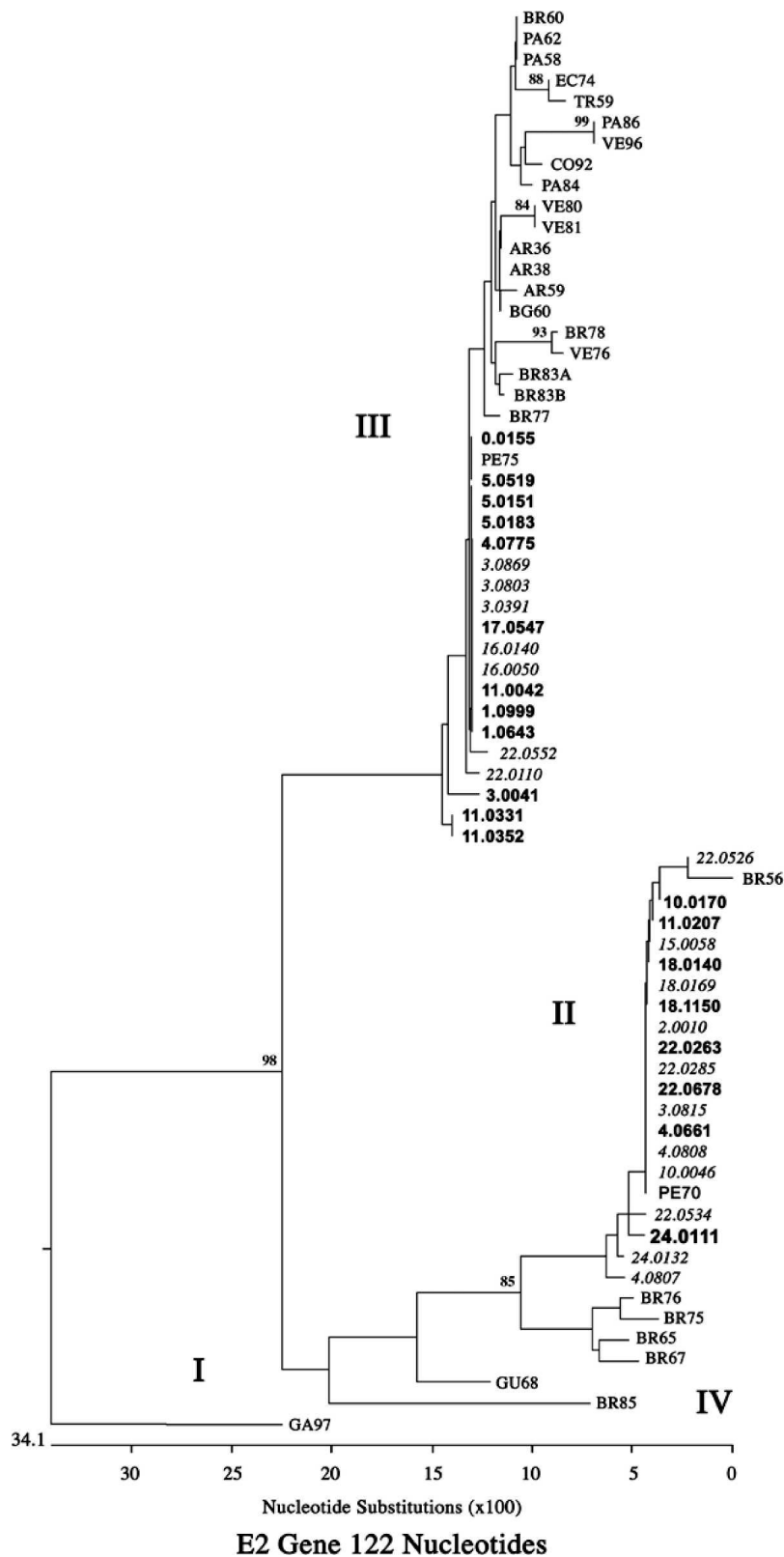† Lineage I = GA97; Lineage II = PE-3.0815; Lineage III = PE-0.0155.

FIGURE 5. Phylogenetic tree of eastern equine encephalomyelitis virus isolates based on a 122-basepair fragment of the envelope 2 (E2) protein gene. Isolates in **bold** and not preceded by letters are from the current study and were isolated in Peru from 1996 to 2000. The remaining viruses are from Brault and others.[5] Bootstrap values ≥ 75 are shown at nodes.

We compared the identity of the two viruses for which we sequenced the entire genome, PE-3.0815 (Lineage II), and PE-0.0155 (Lineage III) with each other and with that of a lineage I isolate (GA97). All three viruses are distinct, with no more than 81.8% identity on the entire genome or 82.1% and 67.5% identity on the E2 or nsP3 genes, respectively (Table 4).

DISCUSSION

An arbovirology study conducted in the Amazon Basin region of Peru afforded us a unique opportunity to study the diversity of members of the genera *Alphavirus*. In particular, the multiple isolations of EEEV allowed us to examine the genomic diversity of these viruses and to compare their relationship with other known virus isolates made at different times and at different geographic locations.

The nsP3 and the E2 genes were selected for sequence and phylogenetic analysis for three reasons. First, sequencing a gene from both the non-structural and the structural regions allowed us to look for possible recombinant viruses, e.g., WEE-like viruses.[10] Although no recombinants were detected, it was possible to identify other unique features within each lineage for the genes selected. Second, the nsP3 gene was chosen because it is comprised of an N-terminal portion that is highly conserved among alphaviruses and a C-terminal portion that is not conserved and varies in both sequence and length.[11] It was a fortuitous choice because the analysis of the nsP3 gene for lineage III strongly suggested (bootstrap values = 100) that the 19 isolations represented in the lineage could be divided into two subclades. And third, the E2 gene was selected because changes in the E2 glycoprotein could result in immunologic differences between the virus isolates. Analysis of the E2 gene also indicated that the lineage III viruses could be divided into the same two subclades (bootstrap values = 85).

The results of the sequencing and phylogenetic analysis of the 38 EEEV isolations suggest that the Lineage II and III viruses co-circulate throughout the year in Peru. Members of both Lineages II and III were collected on the same night at the same study site on several occasions.[2,12] The data also indicate that the distribution between Lineage II and III may vary within a narrow sampling period; more Lineage III virus isolations were obtained during February 1997 compared with more Lineage II virus isolations from February 2000. However, this difference may be an artifact of the relative small number of isolates made during each time period. The data suggest that the best representation of the presence of the two lineages within an area requires sampling throughout the year and over several years.

Because our EEEV isolations were conducted between 1996 and 2001, we were interested in understanding the relationship between these relatively new isolates and those collected during past studies from other regions of South America. To examine the relationship between the EEEV isolates, we conducted a phylogenetic analysis using the sequence data we determined for our virus isolations with those previously published.[5] To accomplish this, we identified and compared a 122-basepair fragment from the E2 gene of our isolates to a corresponding 122-basepair fragment from various South America EEEV published in GenBank.[5] Figure 5 shows a phylogenetic tree that represents the relationship of our current isolates with those analyzed by Brault and others[5] for lineage I, II, III, and IV viruses. One of the results from this comparison is that the samples isolated from Peru in 1970 (Lineage II) and 1975 (Lineage III) were nearly identical to the ones we collected from the same region of Peru, but more than 25 years later, and yet these sequences are distinct from the sequences obtained from viruses isolated from other geographic locations. This suggests, at least for this 122-basepair fragment, that the E2 gene is somewhat conserved in both lineages found in Peru.

Our data suggest that because of the highly conserved nature of EEEV, determining additional sequence data for other regions of the E2 gene, outside the 122-basepair fragment, for the older isolates would enable a better phylogenetic comparison of the viruses. The data in this study also suggest that more than one gene or portion of a gene needs to be sequenced to demonstrate subtle differences in viral structure and thus evolutionary change. The best approach to researching virus evolution would be to sequence the complete genome. Despite the need to sequence more than this 122-basepair fragment, this fragment was able to distinguish between the four lineages and it might be useful in a real-time RT-PCR coupled with microarray technology to not only identify EEEV, but also identify the lineage. The phylogenetic analyses presented here establish a baseline for future work involving the characterization of EEEV isolated from other regions of South and Central America. In addition, the sequence data reported here for the nsP3 and E2 genes may help in understanding the enzootic and epizootic cycles observed for South American and North American EEEV.

Authors' addresses: John P. Kondig and Leonard P. Wasieloski Jr, Diagnostics Systems Division, U.S. Army Medical Research Institute of Infectious Diseases, 1425 Porter Street, Fort Detrick, MD 21702-5011, E-mail: John.Kondig@det.amedd.army.mil. Michael J. Turell, John S. Lee, and Monica L. O'Guinn, Virology Division, U.S. Army Medical Research Institute of Infectious Diseases, 1425 Porter Street, Fort Detrick, MD 21702-5011.

REFERENCES

1. Meegan JM, Digoutte JP, Peters CJ, Shope RE, 1983. Monoclonal antibodies to identify Zinga virus as Rift Valley fever virus. *Lancet 1:* 641.

2. Turell MJ, O'Guinn ML, Jones JW, Sardelis MR, Dohm DJ, Watts DM, Fernandez R, Travassos Da Rosa A, Guzman H, Tesh R, Rossi CA, Ludwig GV, Mangiafico JA, Kondig J, Wasieloski LP Jr, Pecor J, Zyzak M, Schoeler G, Mores CN, Calampa C, Lee JS, Klein TA, 2005. Isolation of viruses from mosquitoes collected in the Amazon Basin region of Peru. *J Med Entomol 42:* 891–898.

3. Morris CD, 1989. Eastern equine encephalitis. Monath TP, ed. *The Arboviruses: Epidemiology and Ecology*. Volume III. Boca Raton, FL: CRC Press, 1–20.

4. Weaver SC, Hagenbaugh A, Bellew LA, Gousset L, Mallampalli V, Holland JJ, Scott TW, 1994. Evolution of alphaviruses in the eastern equine encephalomyelitis complex. *J Virol 68:* 158–169.

5. Brault AC, Powers AM, Chavez CL, Lopez RN, Cachon MF, Gutierrez LF, Kang W, Tesh RB, Shope RE, Weaver SC, 1999. Genetic and antigenic diversity among eastern equine encephalitis viruses from North, Central, and South America. *Am J Trop Med Hyg 61:* 579–586.

6. Lennette EH, Schmidt NJ, eds, 1964. *Diagnostic Procedures for Viral and Rickettsial Diseases: Fluorescent Antibody Techniques*. New York: American Public Health Association, Inc., 177–193.

7. Thompson JD, Higgins DG, Gibson TJ, 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res 22:* 4673–4680.

8. Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD, 2003. Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res 31:* 3497–3500.

9. Howe K, Bateman A, Durbin R, 2002. QuickTree: building huge neighbor-joining trees of protein sequences. *Bioinformatics 18:* 1546–1547.

10. Hahn CS, Lustig S, Strauss EG, Strauss JH, 1988. Western equine encephalitis virus is a recombinant virus. *Proc Natl Acad Sci USA 85:* 5997–6001.

11. LaStarza MW, Lemm JA, Rice CM, 1994. Genetic analysis of the nsP3 region of Sindbis virus: evidence for roles in minus-strand and subgenomic RNA synthesis. *J Virol 68:* 5781–5791.

12. O'Guinn ML, Lee JS, Kondig JP, Fernandez R, Carbajal F, 2004. Field detection of eastern equine encephalitis virus in the Amazon Basin region of Peru using reverse transcription-polymerase chain reaction adapted for field identification of arthropod-borne pathogens. *Am J Trop Med Hyg 70:* 164–171.

13. Kuno G, 1998. Universal diagnostic RT-PCR protocol for arboviruses. *J Virol Methods 72:* 27–41.

14. Martinez HM, 1983. An efficient method for finding repeats in molecular sequences. *Nucleic Acids Res 11:* 4629–4634.

15. Needleman SB, Wunsch CD, 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol 48:* 443–453.